The measurement of composition practice

J. L. Dolby & W. J. Jones

The specification of a computer program for computer-aided typesetting requires that a large number of variables be specified numerically before even the most elementary output can be obtained. A cursory examination of the problem is sufficient to convince most computer men that a careful study will be required if the resulting program is to reproduce standard printing practice. It is, of course, reasonable to argue that any new tool will tend to modify existing practice to a greater or lesser degree and we can expect that computers will have an impact on what is judged to be 'good' composition if they have not already.

However, it is equally true that not all change is for the better and that the wise procedure when considering how to make use of a new piece of equipment is first to make a careful study of existing practice so as to have a firm basis for judging the utility of various proposed 'simplifications'.

This care is particularly necessary when considering the impact of a computer as a piece of operational equipment because of the great flexibility of the computer itself. Potentially, the machine can be used in almost every aspect of the printing trade. At the same time, if the machine is to be used for *any* portion of the business, no matter how small, it is necessary to specify in the computer program what is to be done at a level of detail that far exceeds that normally necessary for communication from one human being to another.

The need for well-documented measurements of composition practice is nicely shown in Barnett's careful documentation of the MIT experiments in computer-aided typesetting. On a number of occasions the author notes that basic program decisions had to be made on a judgement basis with the hope that subsequent users would indicate the need for more appropriate values. However, little has been done to meet this need. Duncan and Resnikoff and Dolby have made extensive studies of the hyphenation and justification portions of the problems, but even this information has not been widely circulated.

As a result it was decided that one of the major objectives of the Computer Typesetting Project at the University of Newcastle-upon-Tyne would be to accumulate a set of measurements on a properly chosen random sample of UK book production to serve as a base for the project programs as well as to be of service to other programming groups in the UK. As the basic data-gathering operation is still under way, this is a preliminary report on the methods of choosing and recording the data and only a brief summary of some of the measurements already analyzed to indicate some of the possible uses that can be made of the complete body of information when it is available.

Sample selection

The first question to be answered is what constitutes the total collection of books printed (and/or published) in the UK. The major listings of UK publications are: *The Bookseller* and the *British National Bibliography,* hereinafter referred to as the BNB. Both publications list the titles of over 20,000 publications each year, and either would have served equally well for the present purpose. The BNB lists every new work published in Great Britain, except for a few classes of publication. *The Bookseller* contains the paperback reprints but is probably not quite so thorough in its coverage of some of the peripheral areas. It was decided to use the BNB as a source definition of the population of books published in Britain and to study the paperback reprints separately at a later time. The 1964 edition of the BNB was used.

A set of two-digit random numbers from Kendall and Smith was selected, the first book to be selected being determined by the first random number, the second by the number produced by adding together the first two random numbers and so on. This procedure provides the required random list in the order in which they are given in the source list and thus reduces the labour of choosing the sample. On the average such a procedure will sample one book in 50, as 50 is the average of all of the numbers between one and 99. A sample of 765 was obtained from a total of 23,575.

Requests for copies were made and as the books were received they were sent to the London College of Printing for measurement. The collection will be maintained at that institution after completion of the study to provide a base for future studies. A separate effort will be necessary to accumulate copies of those books that continue in print and any that go rapidly out of print.

A more significant problem arose in determining what subset of the documents received should be called 'books' as opposed to 'pamphlets'. Some of the documents contained rather special composition problems or contained so little actual composed text as to be trivial from this point of view. As a result it was decided that for the purposes of the preliminary analysis, the following subsets were to be removed for special study at a subsequent date: (a) all documents that were held together in part or in whole by staples (such documents to be tentatively called 'pamphlets'); (b) all children's books with fewer than 100 words to the page on the average; (c) all books printed in a non-latin alphabet; (d) all books printed outside the UK or lithographed from plates made outside the UK; all other documents were declared to be 'books' for the purposes of this study.

Recording problems

Viewed at the level of detail necessary for a computer program, a book—even with the simplifications induced by the restrictive definitions of the preceding section—is a complex of considerable magnitude. The treatment of preliminaries, running text and finals with the many headings, running heads, special lines at chapter beginnings and so on, adds up to a large number of measurements. Potentially, every page of every book may introduce special composition problems and hence special composition practice. As a consequence, some compromise was necessary in choosing what information was to be recorded for each book.

In order to estimate the compromises available to us—particularly in terms of the speed and accuracy with which the figures could be accumulated—a pilot study was conducted at the Newcastle-upon-Tyne School of Printing under the direction of F. R. Marshall. For this some 80-odd books were selected purely on the basis of ease of acquisition and xerographic copies of the first page of each chapter were made. After a preliminary scan of these pages, a form was designed for recording the information contained thereon.

Some findings

On completion the results were analyzed and certain conclusions drawn as to how the information was to be recorded for the main study. The form eventually chosen is shown at the end of this paper. It was found that, on average, each Analysis Form took between two and three hours to complete.

Although several typographical features were predictably common, eg, typeface, line length, page depth, treatment of prelims, a more complex and detailed Analysis Form would have been needed to provide for every variable.

It was found to be important that one person should complete each form. This avoided the temptation to pass on awkward decisions to another person instead of trying to resolve them.

Preliminary analysis

The following is based on the first 165 analysis sheets (which, incidentally, do *not* constitute a random sample of the random sample, and thus may exhibit some bias that will be cleared up when the final results are published).

Type face distribution. One of the problems of interest to those planning computer composition systems is the determination of the number of typefaces the system should be capable of handling as well as the specification of which typefaces should be included in the 'library' for ready access. Estimates of the number of faces in existence vary, but it is generally agreed that the number is large—in the thousands. The statistics collected thus far in this study shed some light on the nature of the problem. As is shown in Table 1, six typefaces tend to dominate the scene, providing almost 75 per cent of the total measured so far:

Table 1 Typeface distribution

Face	Number	Cumulative percentage
Baskerville	39	23.6
Times	29	41.3
Bembo	16	51.0
Plantin	14	59.4
Garamond	13	67.3
Imprint	12	74.6

The next three faces, Century, Fournier, and Juliana account for only 10 more documents with some 25 other faces accounting for the remainder of the sample. With appropriate statistical correction for the sample size we can at the completion of the study summarize by saying that six faces are sufficient for 75 per cent of the books printed in the UK, so many are sufficient for 90 per cent, so many are sufficient for 95 per cent and so on.

Breakdown of Documents by composing machine used. As might be expected for the year 1964, hot metal machinery dominated the total book production. Table 2 provides the breakdown:

Table 2 Composing equipment used

Monotype	116
Linotype	29
Intertype	7
Other	12

Type width distribution. Measurements on line width were made in picas recording to the nearest half pica. Only eight (or 5 per cent) of the books were actually in half pica units, the great majority (92 per cent) coming out in terms of an even number of picas. The nice counting facilities of the computer may well change this situation in the future if it is to be used in cast-off calculations. However, for the moment, it seems reasonable to group the fractional widths with the nearest whole widths to obtain the following result:

Table 3

Type width	Number
21	31
22	29
23	22
24	36
25	13
26	7
27	6
28	4
29	—
30	2
30+	12

Thus over 90 per cent of the books were printed in measures

ranging from 21 picas to 30 picas. The longest measure observed was 46 picas. Other similar statistics show the following type of results:

The most popular number of lines per page is 38 with 90 per cent of the books containing between 30 and 46 lines per page. A curious bias against odd numbers of lines to the page exists: about one and two thirds as many books are produced with even number of lines to the page as opposed to odd number of lines to the page.

The most popular form of treating the first word of a chapter is by the use of caps and small caps. However, there is great diversity here with caps and small caps accounting for just over 20 per cent of the books, with various combinations of caps, small caps, upper and lower case accounting for 72 per cent of the cases.

Almost 90 per cent of the books use chapter headings of one form or another. About 60 per cent have secondary heads, 20 per cent have tertiary heads and only 5 per cent have as many as four headings.

Running heads are used in almost all cases with book title on the verso page and chapter heading on the recto page predominating. However, the reverse situation is not uncommon and almost one quarter of all the books studied used some information other than chapter head and book title.

Summary

Masses of statistics seldom make for interesting reading and no attempt will be made to exhaust the information any further until the complete set of books has been analyzed. However, it is certain that this accumulation of data will be of use to the various groups interested in studying computer typesetting.

Acknowledgements

The authors would like to acknowledge the invaluable help provided them at the beginning of this study by F. R. Marshall of the School of Printing in Newcastle-upon-Tyne, Arthur Phillips of the HMSO, R. C. Gowers of the Publishers' Association, the many members of staff and students of the London College of Printing who contributed their time and their ideas to the data collection portion of this study, and finally to the many publishers who were kind enough to provide the basic material without which the study would have been impossible.

Literature

Barnett, M. P., *Computer typesetting: Experiments & prospects.* Cambridge, Mass, MIT Press, 1965.

Kendall, M. G. & Smith, B. B., *Tracts for Computers, No. XXIV, Tables of Random Sampling Numbers.* Cambridge University Press, Cambridge, 1954.

Wells, A. J. (editor). *The British National Bibliography.* The Council of the British National Bibliography, London.

The Bookseller. J. Whitaker & Sons, London.

UNIVERSITY OF NEWCASTLE UPON TYNE / LONDON COLLEGE OF PRINTING

Computer Typesetting Research Project

Bookwork Format Analysis

GENERAL INSTRUCTIONS

1 Use black (ball point) and do not fold this form.

2 For sections 3, 4, 5, overleaf, ensure that ticks are placed in every horizontal line. Tick NIL if appropriate.

3 Where caps and lower case, or caps and small caps occur together, tick both in appropriate columns.

4 Where several possible answers are given, e.g. Mono/Lino/Inter, strike out those which do not apply.

5 Where indent, flush left, or flush right occur together, tick each in the appropriate column.

6 Special features: use lines at foot of section 2 for any typographical arrangement which is not included elsewhere, e.g. letterspaced heading, unusual indention, etc.

7 When completed, place Analysis Form (unfolded) in book and return to teacher.

SECTION 1—GENERAL INFORMATION

Title:..... Author:.....

Publisher:..... Printer:.....

Date and place of publication:.....

B.N.B. No.....

SECTION 2—TYPOGRAPHICAL SPECIFICATION

Text typeface: Baskerville..... Bembo..... Caledonia.....

Garamond..... Imprint..... Pilgrim.....

Plantin..... Times..... Other.....

Composing system: Mono/Lino/Inter/other.....

Type page area: width.....picas. Lines of text.....

Text columns: single/double..... Measures:.....picas.....points

Running heads: verso page: book title/other.....

recto page: chapter title/other.....

Inter-paragraph leading: NIL/.....points

Chapter opening (spacing)—top of type page to:

Main heading.....picas.....points

Second heading.....picas.....points

Third heading.....picas.....points

Fourth heading.....picas.....points

First text line.....picas.....points

Footnotes.....picas.....points

First paragraph: Initial letter: dropped/raised/nil

If dropped, number of lines.....

Treatment of first words.....

First line of text: full out/indented.....ems

Any special typographical features:.....

.....

.....

.....

.....

DESCRIPTION		Nil	POSITIONING					
			Centre	Flush L	Flush R	Indent	Full out	Other
SECTION 3—PRELIMINARY MATTER								
Half title:	title							
	sub-title							
Title page:	title							
	sub-title							
	author							
	publisher							
	other							
Title page verso:								
Dedication/and Acknowledgement:								
Contents/and illustrations/page:								
	main heading							
	second heading							
	tabular							
	running heads							
Preface/and Foreword/and Introduction:								
	heading							
	running heads							
	semi-display							
	other							
Folios:	top/bottom of page							
SECTION 4—TEXT MATTER								
Chap. opening:	main heading							
	second heading							
	third heading							
	fourth heading							
Text:	headings in text							
	sub-headings in text							
	lists in text							
	captions							
	notes and extracts							
	tables							
	semi-displays							
Folios:	top/bottom of page							
SECTION 5—ODDMENT PAGES								
Appendix:	heading							
	text							
Glossary/and Bib.:	heading							
	text							
Index:	heading							
	text							
Other:	heading							
	text							
Folios:	top/bottom of page							

STYLE							TYPEFACE		TYPE SIZE	COMMENTS
Caps	S. Caps	l.c.	Figs.	Rom.	Ital.	Bold	Text	Other		
										Normal paragraph indent:..................ems
										No. of columns: 1/2/3
										Measure:...........ems...........points
										No. of columns: 1/2/3
										Measure:...........ems...........points

SECTION 6—OTHER INFORMATION
(This section to be completed at a later date)

Page size:.....mm wide x .....mm deep

Paper finish: near white/coloured/coated or smooth/matt or rough

Number of pages: total in book.....

of text (excluding prelims and oddments) but including pages with run-round or inset illustrations.....

Text colour: black/other

Printing process: letterpress/lithography/photogravure/other

Pages and lines	Full out	No. of word spaces	Size of first space	
Page no.:..... line no.:..... No. of para. endings:..... No. of para. openings:.....	Yes/No			
Page no..... line no.:..... No. of para. endings:..... No. of para. openings:.....	Yes/No			
Page no.:..... line no.:..... No. of para. endings:..... No. of para. openings:.....	Yes/No			
Page no.:..... line no.:..... No. of para. endings:..... No. of para. openings:.....	Yes/No			
Page no.:..... line no.:..... No. of para. endings:..... No. of para. openings:.....	Yes/No			

BINDING

No. of pages per section..... Method of binding: sewn/unsewn

No. of sections..... hard (cloth)/soft (paper)

cut flush/turned-in

ILLUSTRATIONS

No. of line.....inset in text.....run-round.....separate pages.....B&W/Colour

No. of halftone.....inset in text.....run-round.....separate pages.....B&W/Colour

grouped.....separate pages.....

PUBLICATION DETAILS

No. of edition:..... Reprint: Yes/No. In same format: Yes/No

Price range:..... Size of edition or printing:.....